RNN with Russell's Circumplex Model for Emotion Estimation and Emotional Gesture Generation

Takuya Tsujimoto, Yasutake Takahashi, and Shouhei Takeuchi
University of Fukui
Bunkyo 3-9-1, Fukui, Fukui, 910-8507, Japan
Email: {ttsujimoto,yasutake,stakeuchi}@ir.his.u-fukui.ac.jp

Yoichiro Maeda
Institute of Technologists
333 Maeya, Gyoda-city, Saitama 361-0038, Japan
Email: maeda@iot.ac.jp

***Abstract*—Interactive Emotion Communication (IEC) has been proposed[1] and studied so far. IEC consists of three processes, recognition of human emotion, generation of robot emotion, and expression of robot emotion. Conventional studies designed those processes by hand one by one. This report proposes a comprehensive system that learns human emotion recognition and robot emotion expression both. The proposed system is a recurrent neural network introducing Russell's circumplex model explicitly and learns human emotion and corresponding motion pattern simultaneously. We show the validity of the proposed method through experiments.**

I. Introduction

In recent years, many interactive robots have been developed and studied in fields such as nursing care and robotic pets. People have more and more opportunities to have interactions with robots these days. Improvement of communication skills of robots is required so that human communicates with a robot naturally and smoothly in the environment where humans and robots live together. In general, people use verbal and nonverbal communication in daily life. They say nonverbal communications dominate over 90% information for the emotional message. Emotional communication between a human and a robot should become more important for natural human-robot interaction in future.

We aim at bidirectional communication based on emotional motion so-called "Interactive Emotion CommunicationIEC."[1] There are various kinds of nonverbal communication, that is, face expression, eye sign, voice pitch, gesture, and so on. We focus on human and humanoid robot gesture in this paper as a nonverbal communication in this paper. IEC is composed of three processes, that is, "human emotion recognition", "expressed emotion decision" and "emotional gesture expression." These processes enable an interactive robot to communicate with human bidirectionally in an emotional way to raise personal affinity of the robot to the human.

Conventional studies on emotion recognition and emotion expression are often studied independently. Oyama and Narita proposed an emotion recognition system based on human facial expression[2]. Taki and Maeda[1] proposed "Fuzzy Emotion Inference System (FEIS) that focuses on the process of "human emotion recognition" by analyzing the human body gesture based on Laban's theory[3], measuring the basic psychological value by fuzzy reasoning and inferring emotion by applying Russell's circumplex model[4]. They also developed emotional gesture generation system and their robot shows an emotional gesture based on the inferred human emotion. However, the emotional gesture is independent from the emotion inference system and designed by hand. Bruce et al. [5] developed a virtual facial emotion expression on a robot. Kanoh et al. developed a real communication robot that focus on emotional facial expression[6]. Itoh et al. developed a facial expression system for a real humanoid robot[7]. Their design of emotion expression might be inspired by actual human emotion expression, however, it is not explicit.

Emotion recognition by gesture and emotion motion generation should be related tightly because a human estimates emotion of other based on his/her own emotional expression. To the best of our knowledge, there are very few references regarding the learning system of recognition and expression of emotion by a gesture.

This paper proposes a novel integrated learning system, "Recurrent Neural Network with Russell's Circumplex Model (RNNRCM)" - which introduces Russell's circumplex model to a Recurrent Neural Network that learns human emotion inference through gesture and robot emotional gesture expression bidirectionally. The RNNRCM realizes the process of "human emotion recognition" and "emotional gesture expression" in the IEC. We confirm the validity of the proposed method with experiments.

II. Interactive Emotion Communication (IEC)

Figure 1 shows the concept of IEC. There are three processes, "(a) human emotion recognition", "(b) expressed emotion decision" and "(c) emotional gesture expression." Process "(a) Human emotion recognition" indicates that a robot recognizes the emotion of the human based on the gesture of the human. The robot decides what kind of emotional gesture should be expressed for rich communication with the human in the process "(b) expressed emotion decision." Then, the robot shows a gesture that expresses the emotion to the human in the process "(c) emotional gesture expression."

III. Recurrent Neural Network with Russell's Circumplex Model (RNNRCM)

Emotional gesture shown by a human is sequential data. A learning model of processes "human emotion recognition" and

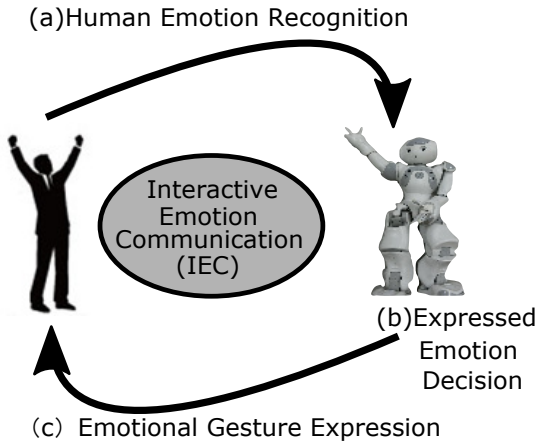


Fig. 1. Concept of IEC

"emotional gesture expression" should handle the sequential data. Recurrent neural network (RNN) is one of the good solutions for the issue because it has an ability of sequential data processing and generation. Tani et al.[8] proposed RNNPB that introduces parametric bias (PB) layer into an RNN. RNNPB can generate robot motions and symbolic values both. Actually, we have tried to use RNNPB for emotional gesture generation and emotion recognition. We suppose that PB would represent a desirable emotion space. We conducted some experiments in which RNNPB learns observed gestures and categorizes them based on PB, however, we found that the RNNPB fails them, unfortunately. We guess the reason why RNNPB is not suitable for the purpose is that there are too many degrees of freedom to configure a desirable emotion space with PB from the observed human emotional gesture. The reasoning suggests us to introduce a certain emotion space to the RNN explicitly.

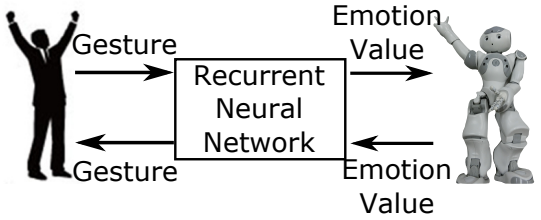


Fig. 2. Concept of RNNRCM

We propose to introduce a certain emotion space into RNN for emotion estimation and emotional gesture generation. Figure 2 shows the concept of our approach. One RNN works for emotion estimation (right arrow) and emotional gesture generation (left arrow) both. A robot observes a human emotional gesture using a simple motion capture system, Microsft's KINECT. The motion capture system provides a sequence data of joint positions of the human. The RNN receives the sequence data of the human joint position, then, output emotion value in the emotion space. The emotion value represents the emotion of the human who demonstrate the gesture. On the other hand, the robot inputs a certain emotion value into the RNN, then, the RNN outputs a sequence data of joint positions of the robot. The robot shows the gesture based on the sequence data of joint position to the opponent human. The former procedure indicates "human emotion recognition" and the latter is "emotional gesture expression" in IEC.

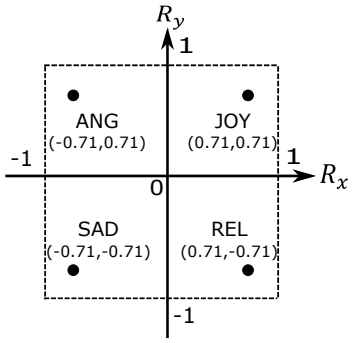


Fig. 3. 4 typical emotion values in motion space based on Russell's circumplex model

We adopt Russell's circumplex model for the emotion space representation. Russell's circumplex model[4] is widely used for emotion expression. It represents emotion on a plane defined by two dimensions. One dimension indicates "arousing and sleepy" (R_x in Fig.3) and the other is "pleasure and unpleasure" (R_y in Fig.3). There are 4 points in the figure each of which represents an emotion value. For example, the upper left point (-0.707, 0.707) represents a typical emotion "anger (ANG)", and the lower right point (0.707, -0.707) represents a typical emotion "relax (REL)." Possible emotion is represented by an emotion value on the plane based on the Russell's circumplex model. The emotion value, represented by R_x and R_y, is introduced into an RNN.

Figure 4 shows the concrete structure of the proposed recurrent neural network with Russell's circumplex model (RNNRCM). The top set of two neurons that handles the emotion value (R_x, R_y) or estimated emotion value $(\hat{R}_x, \hat{R}_y)$ is called emotion layer. P_t^i and V_t^i indicate position and velocity of the ith joint of the human at time t on camera image captured by the simple motion capture system. n indicates the number of joints. Those values are regulated into the range [-1, 1] in advance. $\hat{P}_{t+1}^i$ and $\hat{V}_{t+1}^i$ indicate estimated the position and velocity of the joint at next time step. Figure 5 shows an example of gestures of emotion (a) "joy", (b) "angry", (c) "sad", and (d) "relax." The last set of neurons is a context layer $C_t^1, \cdots, C_t^o$ that feed back the output to the input at next step. The context layer enables the RNN to handle a time series of a gesture.

We prepare emotional gestures corresponding to 4 primitive emotions. RNNRCM learns the emotional gestures and corresponding emotion values simultaneously based on BPTT (Back Propagation Through Time)[9]. The learning data set is composed of four emotional gestures each of which is

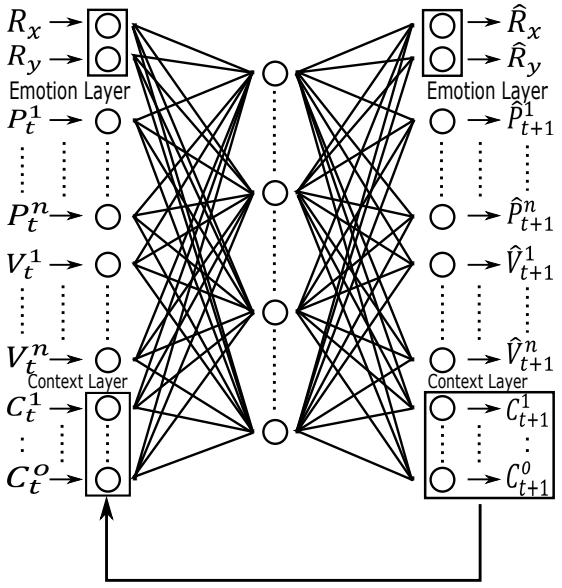
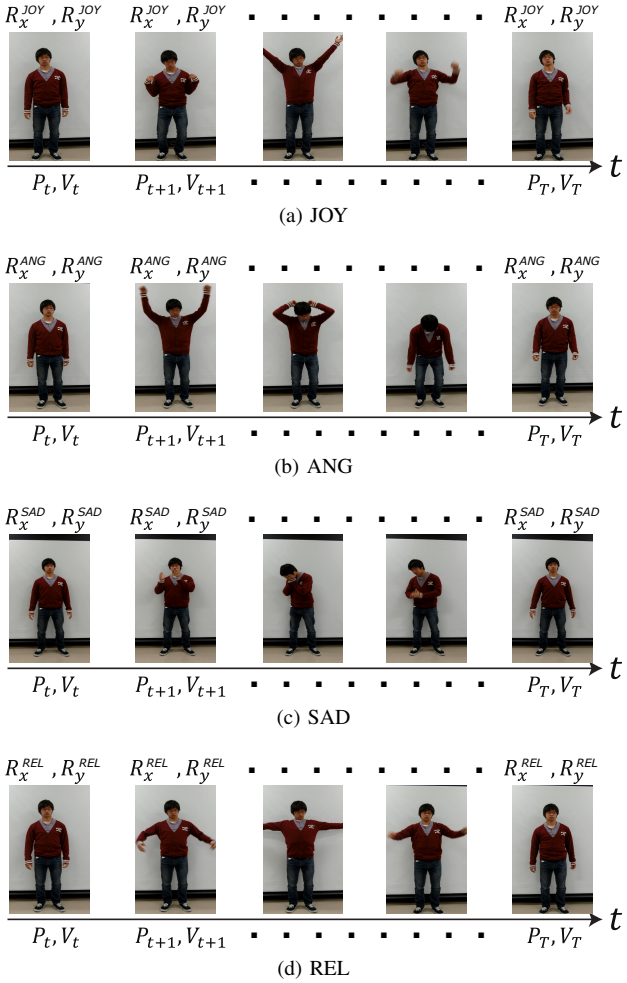


Fig. 4. Concrete Structure of RNNRCM

corresponding to one primitive emotion. During the learning of the RNNRCM, one of the sets of the emotional gesture and the corresponding emotion value is selected in random order and fed to the RNNRCM.

After the learning, the RNNRCM calculates emotion value based on given emotional gesture. Initial values of neurons at the emotion and context layers are set to 0. Position and velocity of the joints at time step 0, P_0^i and V_0^i, is set to the RNNRCM and it generates estimated emotion value $(\hat{R}_x, \hat{R}_y)$. In next step, the estimated emotion value is fed back to the input of the emotion layer and the position and velocity of the joints at time step 1, P_1^i and V_1^i, are set. It repeats these procedures until the end of the given emotional gesture so that the emotion value $(\hat{R}_x, \hat{R}_y)$ is updated accordingly.

The learned RNNRCM can produce an emotional gesture according to the given emotion value. The given emotion value (R_x, R_y) is set to the input neurons at emotion layer and the emotion value is fixed though the corresponding emotional gesture generation. The values at context layer are initialized to 0 at time step 0. The position and velocity of the joints at time step 0, P_0^i and V_0^i, are set to neutral ones. Then, it generates the estimated position and velocity of the joints at time step 1, $\hat{P}_1^i$ and $\hat{V}_1^i$. In next step, the estimated position and velocity of joints are fed back to the input of the position and velocity neurons. The outputs of the context layer are also fed back to the inputs of the context layer. It repeats these procedures for a while so that it generates a learned emotional gesture. A generated emotional gesture is represented as a sequence of the human joints, P_t^i and V_t^i. A particle-filter-based posture mimicking method for a humanoid robot[10] can be applied to acquire an appropriate joint angles of an imitating humanoid robot from the human motion. The method handles the difference of degree of freedom between a human demonstrator and a imitating humanoid robot, accordingly.



(a) JOY

(b) ANG

(c) SAD

(d) REL

Fig. 5. Four typical gestures representing emotions: (a) "joy", (b) "angry", (c) "sad", and (d) "relax"

IV. Experimental Setup

TABLE I shows typical two configurations of RNNRCM for experiments. The parameters are set up by designer's intuition so that there is room to tune them. We have examined other set of parameters for the experiments but we omit them because of the limit of space. Condition 1 in TABLE I is less neurons at the hidden layer and the termination condition is stricter than Condition 2. In general, an RNN with less neurons at the hidden layer learns faster than one with more neurons so that an RNN with less neurons is better. On the other hand, an RNN with more neurons at hidden layer learns more attractors in the network so that the RNNRCM is supposed to maintain more precise emotional gestures.

One male and one female students in their early twenties demonstrate four typical emotional gestures one by one in front of a simple motion capture system, Microsoft's Kinect. The four typical emotion are "joy (JOY)", "anger (ANG)", "sad (SAD)", and "relax (REL)" each of which emotion value is depicted in Figure 3. Each emotional gesture is

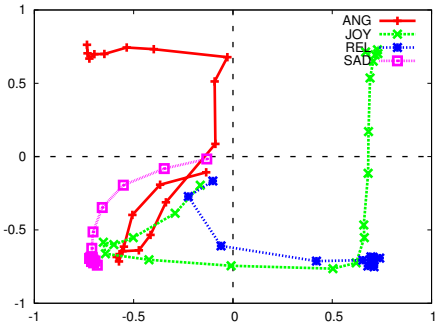
TABLE I
PARAMETERS OF RNNRCM

	Condition 1	Condition 2
Number of Neurons at Hidden Layer	10	100
Number of Neurons at Context Layer	10	10
Learning Rate	0.01	0.01
Learning Termination Condition (Error)	$\leq$0.005	$\leq$ 0.01

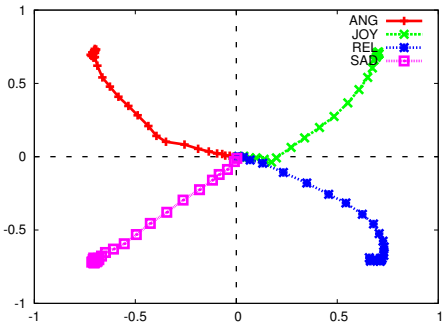
demonstrated within about 2 seconds. The recorded emotional gestures are fed to the RNNRCM for learning.

V. HUMAN EMOTION RECOGNITION BY GESTURE

Human emotion recognition by RNNRCM is evaluated in this section. The two students who are same to the experimental setup demonstrate 4 typical emotional gestures, again. The gestures are not exactly same with the ones for the learning of RNNRCM but they are similar. By showing the demonstarated emotional gesture to the RNNRCM, it calculates the estimated emotion value $(\hat{R}_x, \hat{R}_y)$ step by step.



(a) Condition 1

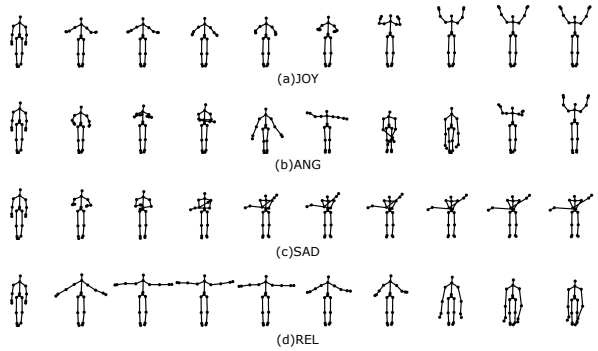


(b) Condition 2

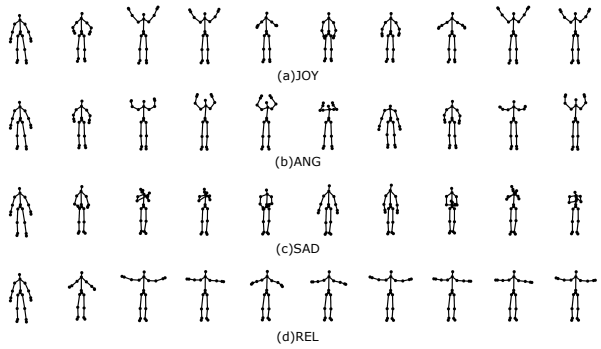
Fig. 6. Sequence of Human Emotion Recognition for 4 Typical Emotional Gesture

Figure 6 shows the results of the experiments. Figure 6(a) shows the emotion recognition results under Condition 1. It shows the recognition is unstable so that all gestures lead to the SAD region at first though the end of the recognition reache to the correct emotion areas. Figure 6(b) shows the emotion recognition results under Condition 2. The RNNRCM shows stable recognition of the emotion. The reason that the RNNRCM under Condition 2 is better than one under Condition 1 is the number of neurons at hidden layer. 10 neurons at the hidden layer are too less to recognize the emotional gestures. 100 neurons seem to be adequate for this experiments.

VI. EMOTIONAL GESTURE GENERATION



(a) Condition 1



(b) Condition 2

Fig. 7. Sequence of Emotional Gesture Generation for 4 Typical Emotions

In this section, RNNRCM ability of emotional gesture generation is evaluated. Emotion value (R_x, R_y) is set to one of 4 typical emotion values shown in Figure 3. The values at context layer are initialized to 0 and the position and velocity of the joints at the time are set to ones of neutral posture. Then, it generates the estimated position and velocity of the joints at consequent time steps.

Figure 7 shows the generated gestures according to the 4 typical emotion values. Figure 7(a) shows the case of Condition 1. RNNRCM generates fairly good gestures for the given emotions, however, the position of the joints sometimes goes wrong so that impossible postures are generated especially in the latter of gesture. For example, the "sad" gesture shows the hand position is too far to realize the posture.

Figure 7(b) shows the case of Condition 2. It shows more realistic gestures according to the given motions than the case of Condition 1. There is no explicit physical constraint of the human body in the proposed RNNRCM. Nevertheless, a large number of neurons at hidden layer seems to take the physical constraint of the human body into consideration.

VII. CONCLUSION

This paper proposed a novel recurrent neural network introducing Russell's circumplex model (RNNRCM) explicitly

that learns human emotion recognition and robot emotion expression both. Demonstrated gestures by human according to typical emotions are learned by the RNNRCM. The RNNRCM recognizes emotion by watching gesture demonstrated by a human and generates emotional gestures according to given emotion. We showed the validity of the proposed method through experiments.

As future work, we are planning to extend the RNNRCM to increase the patterns of each gesture and number of subjects. The current RNNRCM can handle only one pattern of gesture according to each emotion. The recurrent neural network itself has the ability to contain multiple attractors, however, the attractors tend to be unstable if the number of attractors increases, unfortunately.

Another future work is an implementation of the RNNRCM to a real humanoid robot. The current RNNRCM just shows the joint positions of a humanoid robot at 2D coronal plane. It is going to be extended to handle 3D positions of joints of a humanoid robot to generate the emotional gestures in real time. As mentioned at the last paragraph in Section III, we are planning to apply a particle-filter-based posture mimicking approach for a humanoid robot [10].

Last but not least, it is important for us to evaluate the RNNRCM in the context of interactive emotional communication between a human and a humanoid robot. The proposed RNNRCM has an ability to work for "human emotion recognition" and "emotional gesture expression" in real time because the recurrent neural network after learning needs a small amount of calculation. The particle-filter-based posture imitation method for a humanoid robot can be applied to provide appropriate joint angles of the humanoid robot for the generated emotional motion in real time, too. We will verify the ideas through real robot experiments.

REFERENCES

[1] Y. Maeda and R. Taki, "Interactive emotion communication between human and robot," *International Journal of Innovative Computing, Information and Control*, vol. 7, no. 5(B), pp. 2961–2970, 2011.

[2] Y. Oyama and Y. Narita, "A proposal for automatic analysis of emotions using facial charts," *International Journal of Innovative Computing, Information and Control*, vol. 5, no. 3, pp. 717–724, 2009.

[3] R. Laban, *The Mastery of Movement*. Plays, Inc.

[4] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[5] A. Bruce, I. Nourbakhsh, and R. Simmons, "The role of expressiveness and attention in human-robot interaction," in *Proceedings of 2002 IEEE International Conference on Robotics and Automation (ICRA'02)*, vol. 4, 2002, pp. 4138–4142.

[6] M. Kanoh, S. Iwata, S. Kato, and H. Itoh, "Emotive facial expressions of sensitivity communication robot 'ifbot'," *Kansei Engineering International*, vol. 5, no. 3, pp. 35–42, 2005.

[7] K. Itoh, H. Miwa, M. Matsumoto, M. Zecca, H. Takanobu, S. Rocdella, M. C. Carrozza, P. Dario, and A. Takanishi, "Various emotion expressions with emotion expression humanoid robot," in *Proceeding of the 1st IEEE Technical Exhibition Based Conference on Robotics and Automation*, 2004, pp. 35–36.

[8] J. Tani, M. Ito, and Y. Sugita, "Self-organization of distributedly multiple behavior schemata in a mirror system: Reviews of robot experiments using rnnpb," *Neural Networks*, vol. 24, no. 5, pp. 1273–1289, 2004.

[9] R.J.Williams and J.Peng, "An efficient gradient-based algorithm for online training of recurrent network trajectories," *Neural Computation*, pp. 490–501, 1990.

[10] Y. TAKAHASHI and K. SAKAKIBARA, "Real-time joint angle estimation for a humanoid robot to imitate human motion using particle filter (in japanese)," in *JSAI Technical report, SIG-Challenge 042*, May 2015, pp. 21–23.